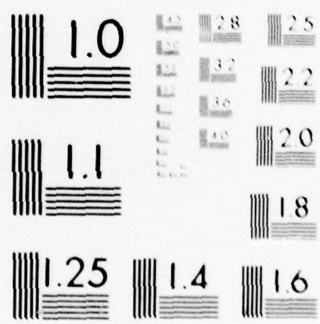
AD-A074 223 PITTSBURGH UNIV PA INST FOR STATISTICS AND APPLICATIONS F/G 12/2
ESTIMATION OF VARIANCE COMPONENTS.(U)
JUL 79 C R RAO, J KLEFFE F49620-79-C-0161
UNCLASSIFIED TR-79-1 AFOSR-TR-79-0974 NL

| OF |

AD A074223

END
DATE
FILMED
10-79
DDC


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6

Institute for Statistics and Applications

University of Pittsburgh

ESTIMATION OF VARIANCE COMPONENTS

C. Radhakrishna Rao*

University of Pittsburgh

and

Jurgen Kleffe

Akademie der Wissenschaften der D.D.R.

July 1979
Technical Report No. 79-1

Institute for Statistics and Applications
Department of Mathematics and Statistics
University of Pittsburgh

*The work of this author is sponsored by the Air Force Office of Scientific Research, Air Force Systems Command under Contract F49620-79-C-0161. Reproduction in whole or in part is permitted for any purpose of the United States Government.

(9) Interim rept., (11) Jul 79

(6)

ESTIMATION OF VARIANCE COMPONENTS

(12) 63p.

(10) C. Radhakrishna Rao*

University of Pittsburgh

~~and~~

Jurgen Kleffe

Akademie der Wissenschaften der D.D.R.

(14) TR-79-1

(16) 2304

(17) A5

(18) AFOSR

(19) TR-79-0974

*The work of this author is sponsored by the Air Force Office of Scientific Research, Air Force Systems Command under Contract (15) F49620-79-C-0161. Reproduction in whole or in part is permitted for any purpose of the United States Government.

411 385

1/B

1. INTRODUCTION

The usual mixed linear model discussed in the literature on variance components is

$$Y = X\beta + U_1\phi_1 + \ldots + U_p\phi_p + \varepsilon \tag{1.1}$$

where $X, U_1, \ldots, U_p$ are known matrices, β is a fixed unknown vector parameter and $\phi_1, \ldots, \phi_p$, ε are unobservable random variables (r.v.'s) such that

$$E(\varepsilon) = 0, \quad E(\phi_i) = 0, \quad E(\phi_i\phi_j') = 0, \quad i \neq j, \quad E(\varepsilon\phi_i') = 0$$

$$E(\varepsilon\varepsilon') = \sigma_o^2 I_n, \quad E(\phi_i\phi_i') = \sigma_i^2 I_{n_i}. \tag{1.2}$$

The unknown parameters $\sigma_o^2, \sigma_1^2, \ldots, \sigma_p^2$ are called variance components.

Some of the early uses of such models are due to Yates and Zacopancy (1935) and Cochran (1939) in survey sampling, Yates (1940) and Rao (1947, 1956) in combining intra and interblock information in design of experiments, Fairfield Smith (1936), Henderson (1950), Panse (1946) and Rao (1953) in the construction of selection indices in genetics, and Brownlee (1953) in industrial applications. A systematic study of the estimation of variance components was undertaken by Henderson (1953) who proposed three methods of estimation.

The general approach in all these papers was to obtain $p+1$ quadratic functions of Y, say $Y'Q_iY$, $i = 1, \ldots, p+1$, which are invariant for translation of Y by $X\alpha$ where α is arbitrary, and solve the equations

$$Y'Q_iY = E(Y'Q_iY) = a_{io}\sigma_o^2 + a_{i1}\sigma_1^2 + \ldots + a_{ip}\sigma_p^2 \tag{1.3}$$

$$i = 0,1,\ldots,p.$$

The method of choosing the quadratic forms was intuitive in nature (see Henderson, 1953) and did not depend on any stated criteria of estimation. The entries in the ANOVA table giving the sums of squares due to different effects were considered as good choices of the quadratic forms in general. The ANOVA technique provides good estimators in what are called balanced designs (see R. L. Anderson, 1975 and R. L. Anderson and P. P. Crump, 1967) but, as shown by Seely (1975) such estimators may be inefficient in more general linear models. For a general discussion of Henderson's methods and their advantages (computational simplicity) and limitations (lack of uniqueness, inapplicability and inefficiency in special cases) the reader is referred to papers by Searle (1968, 1971), Seely (1975), Olsen, Sealy and Birkes (1976) and Harville (1977, p. 335).

A completely different approach is the ML (maximum likelihood) method initiated by Hartley and Rao (1967). They considered the likelihood of the unknown parameters $\beta, \sigma_o^2, \ldots, \sigma_p^2$ based on observed Y and obtained the likelihood equations by computing the derivatives of likelihood with respect to the parameters. Patterson and Thompson (1975) considered the marginal likelihood based on the maximal invariant of Y, i.e., only on $B'Y$ where B is a matrix orthogonal to X and obtained what are called restricted maximum likelihood (RML) equations. Harville (1977) has given a review of the ML and RML methods and the computational algorithms associated with them.

One criticism of the ML estimators is that they may be heavily biased so that some caution is needed when they are used as estimates of individual parameters for taking decisions or for using them in the place of true values to obtain an efficient estimate of β. The problem is not accute if the exact distribution of the ML estimators is known, since in that case appropriate adjustments can be made in the individual estimators before using them. The general large sample properties associated with ML estimators are misleading in the absence of studies on the orders of sample sizes for which these properties hold in particular cases.

The bias in RML estimators may not be large even in small samples. But both ML and RML estimators are functions of $B'Y$, the maximal invariant of Y, and there are important practical cases where reduction of Y to $B'Y$ results in non-identifiability of individual parameters, in which case neither ML nor RML is applicable. The details are given in Section 5.

Rao (1970, 1971a, 1971b, 1972, 1973) proposed a general method called MINQE (Minimum norm quadratic estimation) the scope of which has been extended to cover a variety of situations by Focke and Dewess (1972), Kleffe (1975, 1976, 1977a,b, 1978, 1979), J. N. K. Rao (1973), Fuller and Rao (1978), Poduri Rao and Chaubey (1978), Pukelsheim (1977, 1978), Sinha and Wieand (1977) and Rao (1979). The method is applicable to a general linear model

$$Y = X\beta + \epsilon, \quad E(\epsilon\epsilon') = \theta_1 V_1 + \ldots + \theta_p V_p \tag{1.4}$$

where no structure need be imposed on ϵ and no restrictions are placed on θ_i or V_i. (In the model (1.1), $\theta_i \geq 0$ and V_i are non-negative definite).

In the MINQE theory, we define what is called a natural estimator of a linear function $f'\theta$ of θ in terms of the unobservable r.v. ε in (1.4), say $\varepsilon' N \varepsilon$. Then the estimator $Y'AY$ in terms of the observable r.v. Y is obtained by minimizing the norm of the difference between the quadratic forms $\varepsilon' N \varepsilon$ and $Y'AY = (X\beta + \varepsilon)'A(X\beta + \varepsilon)$. The universality of the MINQE method arises from the following observations:

(a) It offers a wide scope in the choice of the norm depending on the nature of the model and prior information available.

(b) One or more restrictions such as invariance, unbiasedness and non-negative definiteness can be placed on $Y'AY$ depending on the desired properties of the estimators.

(c) The method is applicable is situations where ML and RML fail.

(d) There is an automatic provision for incorporating available prior information on the unknown parameters β and θ.

(e) Further, ML and RML estimators can be exhibited as iterated versions of suitably chosen MINQE's.

(f) The MINQE equation provides a natural numerical algorithm for computing the ML or RML estimator.

(g) For a suitable choice of the norm, the MINQ estimators provide minimum variance estimators of θ when Y is normally distributed.

It has been mentioned by some reviewers of the MINQE theory that the computations needed for obtaining the MINQ estimators are somewhat heavy. It is true that the closed form expressions given for MINQE's contain inverses of large order matrices, but they can be computed in a simple way in special cases that arise in practice. The computations in such cases are of the same order of magnitude as obtaining sums of squares in the ANOVA table appropriate for the linear model. Perhaps, some research is needed in developing simple numerical algorithms in more complicated cases. It is certainly not true that the computation of MLE or RMLE is simpler than that of MINQE. Both may have the same order of complexity in the general case.

2. MODELS OF VARIANCE AND COVARIANCE COMPONENTS

2.1 General Model

There is a large variety of models of variance and covariance components used in research work in biological and behavioral sciences. They can all be considered in a unified frame work under a general Gauss-Markoff (GM) model

$$Y = X\beta + \varepsilon \tag{2.1.1}$$

where Y is n-vector random variable, X is $n \times m$ matrix, β is m-vector parameter and ε is n-vector variable. The models differ mainly in the structure imposed on ε. The most general formulation is

$$E(\varepsilon) = 0 \tag{2.1.2}$$

$$D(\varepsilon) = \theta_1 V_1 + \ldots + \theta_p V_p = V(\theta) = V_\theta \tag{2.1.3}$$

where D stands for the dispersion (variance covariance) matrix, $\theta' = (\theta_1, \ldots, \theta_p)$ is unknown vector parameter and $V_1, \ldots, V_p$ are known symmetric matrices. We let $\beta \varepsilon R^m$ and $\theta \varepsilon F$ (open set) $\subset R^p$ such that $V(\theta) \geq 0$ (i.e., nonnegative definite). In the representation (2.1.3) we have not imposed any restriction such as $\theta_i \geq 0$ or V_i is nonnegative definite.

It may be noted that any arbitrary $n \times n$ dispersion matrix $\Theta = (\theta_{ij})$ can be written in the form (2.1.3)

$$\Sigma\Sigma\theta_{ij} V_{ij} \tag{2.1.4}$$

involving a maximum of $p = n(n-1)/2$ unknown parameters θ_{ij} and known matrices V_{ij}, but in models of practical interest p has a relatively small value compared to n.

2.2 Variance Components

A special case of the variance components model is when ε has the structure

$$\varepsilon = U_1\phi_1+\ldots+U_p\phi_p \tag{2.2.1}$$

where U_i is $n \times m_i$ given matrix and ϕ_i is m_i-vector r.v. such that

$$E(\phi_i) = 0, \quad E(\phi_i\phi_j') = 0 \ \ i\neq j, \quad E(\phi_i\phi_i') = \sigma_i^2 I_{m_i}. \tag{2.2.2}$$

In such a case

$$V(\theta) = \theta_1 V_1+\ldots+\theta_p V_p \tag{2.2.3}$$

where $V_i = U_iU_i' \geq 0$ and $\theta_i = \sigma_i^2 \geq 0$. Most of the models discussed in literature are of the type (2.2.1) leading to (2.2.3).

In the classical regression model $p = 1$. Other examples are one and two way classification models with fixed and random effects.

The complete GM model when ε has the structure (2.2.1) is

$$Y = X\beta + U_1\phi_1+\ldots+U_p\phi_p$$

$$E(\phi_i) = 0; \quad E(\phi_i\phi_j') = 0 \ \ i\neq j; \quad E(\phi_i\phi_i') = \sigma_i^2 I_{m_i}. \tag{2.2.4}$$

The associated statistical problems are:

(a) Estimation of β

(b) Estimation of σ_i^2, $i = 1,\ldots,p$

(c) Estimation of ϕ_i, $i = 1,\ldots,p$. (2.2.5)

The last problem arises in the construction of selection indices in genetics, and some early papers on the subject providing a satisfactory solution are due to Fairfield Smith (1936), Panse (1946) based on an idea suggested by Fisher, and Henderson (1950). A theoretical justification of the method employed by these authors and associated tests of significance are given in Rao (1953).

A particular case of the model (2.2.4) is where it can be broken down into a number of submodels

$$Y_1 = X_1\beta + \varepsilon_1, \ldots, Y_p = X_p\beta + \varepsilon_p \tag{2.2.6}$$

where Y_i is n_i-vector variable and

$$E(\varepsilon_i) = 0, \quad E(\varepsilon_i\varepsilon_i') = \theta_i I_{n_i}, \quad E(\varepsilon_i\varepsilon_j') = 0. \tag{2.2.7}$$

Note that the β parameters are the same in all submodels, and in some situations the design matrices $X_1, \ldots, X_p$ may also be the same. The model (2.2.6) with the covariance structure (2.2.7) is usually referred to as one with "heteroscedastic variances" and the problem of estimating β as that of estimating a "common mean".

2.3 Variance and Covariance Components

We assume the same structure (2.2.1) for ε but with different covariances for the ϕ_i's

$$E(\phi_i) = 0, \quad E(\phi_i\phi_i') = \Lambda_i, \; i = 1, \ldots, k$$

$$E(\phi_i\phi_i') = \sigma_i^2 I_{m_i}, \; i = k+1, \ldots, p \tag{2.3.1}$$

$$E(\phi_i\phi_j') = 0, \quad i \neq j$$

leading to

$$V(\theta) = U_1\Lambda_1U_1'+\ldots+U_k\Lambda_kU_k' + \sigma_{k+1}^2U_{k+1}U_{k+1}'+\ldots+\sigma_p^2U_pU_p' \tag{2.3.2}$$

where $\Lambda_i \geq 0$. In some practical problems Λ_i are all the same and there is only one σ^2 in which case (2.3.1) becomes

$$V(\theta) = U_1\Lambda U_1'+\ldots+U_k\Lambda U_k' + \sigma^2 I. \tag{2.3.3}$$

2.4 Random Regression Coefficients

This is a special case of the variance and covariance components model considered in 2.3 where ε has the structure

$$\varepsilon = X\phi_1 + \phi_2, \quad E(\phi_1\phi_1') = \Lambda, \quad E(\phi_2\phi_2') = \sigma^2 I \tag{2.4.1}$$

the compounding matrix for ϕ_1 being the same as for β leading to the GM model

$$Y = X\beta + X\phi_1 + \phi_2$$

$$D(\varepsilon) = X\Lambda X' + \sigma^2 I. \tag{2.4.2}$$

In general, we have repeated observations on the model (2.4.2) with different X's

$$Y_i = X_i\beta + X_i\phi_{1i} + \phi_{2i}, \quad i = 1,\ldots,t \tag{2.4.3}$$

leading to the model

$$Y = X\beta + \varepsilon \tag{2.4.4}$$

with

$$D(\varepsilon) = \begin{pmatrix} X_1 \Lambda X_1 + \sigma^2 I & & 0 \\ & \ddots & \\ 0 & & X_t \Lambda X_t + \sigma^2 I \end{pmatrix} \tag{2.4.5}$$

all the off diagonal blocks being null matrices. A discussion of such models is contained in Fisk (1967), Rao (1965, 1967), Swamy (1971), and Spjotvoll (1977).

2.5 Intraclass Correlation Model

We shall illustrate an intraclass correlation model with special reference to two way classified data with repeated observations in each cell

$$y_{ijk}, \ i = 1,\dots,p; \ \ j = 1,\dots,q; \ \ k = 1,\dots,r. \tag{2.5.1}$$

We write

$$Y_{ijk} = \mu_{ijk} + \varepsilon_{ijk} \tag{2.5.2}$$

where μ_{ijk} are fixed parameters with a specified structure, and

$$E(\varepsilon_{ijk}) = 0, \ \ E(\varepsilon^2_{ijk}) = \sigma^2$$

$$\begin{aligned} E(\varepsilon_{ijr}\varepsilon_{ijs}) &= \sigma^2\rho_1, \quad r\neq s \\ E(\varepsilon_{ijr}\varepsilon_{iks}) &= \sigma^2\rho_2, \quad j\neq k, \ \ r\neq s \\ E(\varepsilon_{ijr}\varepsilon_{tks}) &= \sigma^2\rho_3, \quad i\neq t, \ \ j\neq k, \ \ r\neq s. \end{aligned} \tag{2.5.3}$$

This dispersion matrix of (Y_{ijk}) can be exhibited in the form (2.1.3) with four parameters $\sigma^2, \rho_1, \rho_2, \rho_3$. A model of the type

(2.5.2) is given in Rao (1973, p. 258).

2.6 Multivariate Model

A k-variate linear model is of the form

$$(Y_1:\ldots:Y_k) = X(\beta_1:\ldots,\beta_k) + (\epsilon_1:\ldots:\epsilon_k)$$

$$E(\epsilon_i\epsilon_j') = \sigma_{ij}^{(1)}V_1+\ldots+\sigma_{ij}^{(p)}V_p. \qquad (2.6.1)$$

Denoting $\overline{Y} = (Y_1',\ldots,Y_k')$, $\overline{\beta} = (\beta_1',\ldots,\beta_k')$, $\overline{\epsilon} = (\epsilon_1',\ldots,\epsilon_k')$, the multivariate model may be written as a univariate model

$$\overline{Y} = (I \otimes X)\overline{\beta} + \overline{\epsilon}$$

$$E(\overline{\epsilon}\overline{\epsilon}') = \sum_{i=1}^{p} (\Theta_i \otimes V_i) \qquad (2.6.2)$$

where Θ_i are $(k\times k)$ matrices of variance and covariance components. In the multivariate regression model $p = 1$, in which case

$$E(\overline{\epsilon}\overline{\epsilon}') = \Theta \otimes V. \qquad (2.6.3)$$

We may specify structures for ϵ analogous to (2.2.1) in the univariate case

$$\epsilon_i = U_1\phi_{1i}+\ldots+U_p\phi_{pi},\ i = 1,\ldots,k$$

$$E(\phi_{im}\phi_{jm}') = \sigma_{ij}^{(m)},\quad E(\phi_{ir}\phi_{jh}') = 0 \quad r\neq s. \qquad (2.6.4)$$

For special choices of U_i, we obtain multivariate one, two,...way mixed models.

Models of the type (2.6.2) have been considered by Krishnaiah and Lee (1974). They discuss methods of estimating the covariance matrices Θ_i and testing the hypothesis that a covariance matrix has the structure (2.6.2).

3. ESTIMABILITY

3.1 Unbiasedness

Let us consider the univariate GM model (2.1.1) with the covariance structure (2.1.3)

$$Y = X\beta + \varepsilon, \quad D(\varepsilon) = \theta_1 V_1 + \ldots + \theta_p V_p \tag{3.1.1}$$

and find the conditions under which linear functions $f'\theta$ can be estimated by functions of Y subject to some constraints. The classes of estimators considered are as follows:

$$\mathcal{Q} = \{Y'AY,\ A \text{ symmetric}\} \tag{3.1.2}$$

$$\mathcal{U}_f = \{g(Y) : E[g(Y)] = f'\theta \ \forall\ \beta \varepsilon R^m,\ \theta \varepsilon F\} \tag{3.1.2}$$

$$\mathcal{I} = \{g(Y) = g(Y + X\alpha)\} \tag{3.1.3}$$

Theorem 3.1.1 provides conditions for unbiased estimability.

<u>Theorem 3.1.1</u>. Let the linear model be as in (3.1.1). Then:

(i) The estimator $Y'AY$ is unbiased for $\gamma = f'\theta$ iff

$$X'AX = 0, \quad \operatorname{tr} AV_i = f_i, \ i = 1,\ldots,p. \tag{3.1.5}$$

(ii) There exists an unbiased estimator $\hat{\gamma} \varepsilon \mathcal{Q}$ iff $f \varepsilon S(H)$,

$$H = (h_{ij}), \quad h_{ij} = \operatorname{tr}(V_i V_j - PV_i PV_j) \tag{3.1.6}$$

where P is the projection operator onto $S(X)$.

(iii) If Y has multivariate normal distribution, then $\mathcal{U}_f \cap \mathcal{Q}$ is not empty.

The results (i) and (ii) are discussed in Seely (1970), Rao (1970, 1971) and Focke and Dowess (1972) and (iii) in Pincus (1974).

Note 1: Result (ii) holds if in (3.1.6) we choose

$$h_{ij} = \mathrm{tr}(V_i(I - P)V_j). \tag{3.1.7}$$

Note 2: In the special case $V_iV_j = 0$ for $i \neq j$, θ_i, the i-th individual parameter, is unbiasedly estimable iff $MV_i \neq 0$ where $M = I - P$.

Lemma 3.1.1. The linear space Γ of all unbiasedly estimable linear functions of θ is

$$\Gamma = \{\underset{\theta}{E}\, Y'AY: A \in sp(V_1 - PV_1P, \ldots, V_p - PV_pP)\} \tag{3.1.8}$$

where $sp(A_1, \ldots, A_p)$ is the set of all linear combinations of $A_1, \ldots, A_p$.

Let us consider the multivariate model (2.6.2) written in a vector form

$$\overline{Y} = (I \otimes X)\overline{\beta} + \overline{\varepsilon}$$

$$E(\overline{\varepsilon}\overline{\varepsilon}') = \Theta_1 \otimes V_1 + \ldots + \Theta_p \otimes V_p \tag{3.1.9}$$

where Θ_i are $k \times k$ matrix variance-covariance components.

Lemma 3.1.2. The parametric function $\gamma = \Sigma f_i \, \mathrm{tr} C\, \Theta_i$ is unbiasedly estimable from the model (3.1.9) iff $f'\theta$ is so from the univariate model (3.1.1).

Lemma 3.1.3. The class Γ of unbiasedly estimable linear functions of elements of Θ_i, $i = 1,\ldots,p$ in (3.1.9), is

$$\Gamma = \{\gamma = \Sigma\, \mathrm{tr} C_i \Theta_i : C_i \text{ are such that } Hb = 0 \Rightarrow \Sigma\, b_i C_i = 0\} \tag{3.1.10}$$

where H is as defined in (3.1.6).

3.2 Invariance

An estimator is said to be invariant for translation of the parameter β in the linear model (3.1.1) if it belongs to the class (3.1.4). Theorem 3.2.1 provides the conditions under which estimators belonging to the class $\mathcal{U}_f \cap \mathcal{I}$ exist.

Theorem 3.2.1. Let the linear model be as in (3.1.1). Then:

(i) The estimator $Y'AY \in \mathcal{U}_f \cap \mathcal{I}$ iff

$$AX = 0, \quad \mathrm{tr}\, AV_i = f_i, \quad i = 1,\ldots,p, \tag{3.2.1}$$

(ii) There exists an unbiased estimator in class $\mathcal{Q} \cap \mathcal{I}$ iff $f \in \mathcal{I}(H_M)$ where

$$H_M = (h_{ij}), \quad h_{ij} = \mathrm{tr}(MV_iMV_j), \quad M = I - P. \tag{3.2.2}$$

(iii) Under the assumption of normality of Y, the result (3.2.2) can be extended to the class $\mathcal{I}$.

<u>Note</u>: In (3.2.2), we can choose

$$h_{ij} = \mathrm{tr}(BB'V_iBB'V_j) \tag{3.2.3}$$

where B is any choice of $X^{\perp}$, i.e., B is a matrix of maximum rank such that $B'X = 0$.

Lemma 3.2.1. The linear space of all invariantly unbiasedly estimable linear functions of θ is

$$\Gamma_I = \{E_\theta\, Y'MAMY: \ A \in sp(V_1 - PV_1P, \ldots, V_p - PV_pP)\}. \tag{3.2.4}$$

Lemma 3.2.2. If $f'\theta$ is invariantly unbiasedly estimable from the model (3.1.1) then so is $\gamma = \Sigma f_i \, \mathrm{tr} C\, \Theta_i$ from the model (3.1.9).

Lemma 3.2.3. All invariantly unbiasedly estimable linear functions of the elements of $\Theta_1, \ldots, \Theta_p$ in the model (3.1.9) belong to the set

$$\Gamma_{UI} = \{\gamma = \Sigma\, \mathrm{tr}\, C_i\, \Theta_i: \ C_i \text{ are such that } H_M b = 0 \Rightarrow \Sigma b_i C_i = 0\}. \tag{3.2.5}$$

Note: We can estimate any member of the class (3.2.5) by functions of the form

$$\hat{\gamma} = \Sigma\, \mathrm{tr}(C_j\, Y'A_jY)$$

where $A_1, \ldots, A_p$ are matrices arising in invariant quadratic unbiased estimation in the univariate model (3.1.1).

3.3 Examples

Consider the model with 4 observations

$$Y_1 = \beta_1 + \epsilon_1, \quad Y_2 = \beta_1 + \epsilon_2, \quad Y_3 = \beta_2 + \epsilon_3, \quad Y_4 = \beta_2 + \epsilon_4$$

where ϵ_i are all uncorrelated and $V(\epsilon_1) = V(\epsilon_3) = \sigma_1^2$ and $V(\epsilon_2) = V(\epsilon_4) = \sigma_2^2$. The matrices X, V_1, V_2 and P the projection operator are easily seen to be

$$X = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}, \quad V_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad V_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix},$$

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

The matrices H and H_M of Theorems 3.1.1 and 3.2.1 are

$$H = \begin{pmatrix} \frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{3}{2} \end{pmatrix}, \quad H_M = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Since H is of full rank, σ_1^2 and σ_2^2 are individually unbiasedly estimable. But H_M is of rank one and the unit vectors do not belong to the space $S(H_M)$ and therefore, σ_1^2 and σ_2^2 are not individually unbiasedly estimable by invariant quadratic forms.

Consider the model $Y = X\beta + X\phi + \varepsilon$ where β is a fixed vector parameter and ϕ is a vector of random effects such that $E(\phi) = 0$, $E(\phi\phi') = \sigma_2^2 I_m$, $E(\phi\varepsilon') = 0$, $E(\varepsilon\varepsilon') = \sigma_1^2 I_n$. Let $Y'AY$ be an unbiased estimate of σ_2^2. Then we must have

$$X'AX = 0, \quad \operatorname{tr} AXX' = 1, \quad \operatorname{tr} A = 0$$

which is not consistent since $X'AX = 0 \Longrightarrow \operatorname{tr} AXX' = 0$. Hence unbiased estimators of σ_2^2 do not exist.

4. MINIMUM VARIANCE UNBIASED ESTIMATION-NORMAL CASE

4.0 Notations

In Section 3, we obtained conditions for unbiased estimability of $f'\theta$ in the linear model

$$Y = X\beta + \varepsilon, \quad D(\varepsilon) = \theta_1 V_1 + \ldots + \theta_p V_p = V_\theta \tag{4.0.1}$$

restricting the class of estimators to quadratic functions of Y. In this section we do not put any restriction on the class of estimators but assume that

$$Y \sim N_n(X\beta, V_\theta), \quad \beta \in R^m, \quad \theta \in F \tag{4.0.2}$$

i.e., n variate normal, and V_θ is p.d. for $\theta \in F$. The condition that V_θ is p.d. is assumed to simplify presentation of results, and is satisfied in many practical situations.

First, we derive the locally minimum variance unbiased estimator (LMVUE) of $f'\theta$ at a chosen point (β_o, θ_o) in $R^m XF$. If the estimator is independent of β_o, θ_o then we have a uniformly minimum variance unbiased estimator (UMVUE). Such estimators do not exist except in simple cases. In the general case we suggest the use of LMVUE with a suitable choice of β_o, θ_o based on previous experience or apriori considerations. We also indicate an iterative method which starts with an initial value (β_o, θ_o), gets an improved set (β_1, θ_1), and provides in the limit an iterated MVE.

LMVUE's are obtained in the class of quadratic estimators by La Motte (1973) under the assumption of normality and by Rao (1971) in the general case. Such estimators were designated by Rao as

MIVQUE (minimum variance quadratic unbiased estimator). In this section we show that, under the normality assumption, MIVQUE is LMVUE in the whole class of unbiased estimators.

4.1 Locally Minimum Variance Unbiased Estimation

Definition 4.1.1. An estimator $\hat{\gamma}_*$ is called LMVUE of its expected value at $(\beta_o,\theta_o) \varepsilon R^m XF$ iff

$$V(\hat{\gamma}_*|\beta_o,\theta_o) \leq V(\hat{\gamma}|\beta_o,\theta_o) \tag{4.1.1}$$

for all $\hat{\gamma}$ such that

$$E(\hat{\gamma}_*) = E(\hat{\gamma}) \forall (\beta,\theta)\varepsilon \; R^m XF. \tag{4.1.2}$$

We use the following notations:

$$V_\theta = \theta_1 V_1 + \ldots + \theta_p V_p$$

$$A_i = V_\theta^{-1}(V_i - P_\theta V_i P_\theta')V_\theta^{-1}, \quad P_\theta = X(X'V_\theta^{-1}X)X'V_\theta^{-1}$$

$$K_\theta = (\text{tr } A_i V_i)$$

$$k_{\beta,\theta} = [(Y-X\beta)'A_1(Y-X\beta),\ldots,(Y-X\beta)'A_p(Y-X\beta)]'. \tag{4.1.3}$$

Let (β_o,θ_o) be an apriori value of (β,θ). Then applying the result (3.1.6) of Theorem 3.1.1 we find that $f'\theta$ is unbiasedly estimable iff

$$f \; \varepsilon \, S(K_{\theta_o}).$$

Theorem 4.1.1 provides an explicit expression for the LMVUE.

Theorem 4.1.1. Let f satisfy the condition (4.1.4) and K_θ, $k_{\beta,\theta}$ be as defined in (4.1.3). Then the LMVUE of $f'\theta$ at (β_o,θ_o) is

$$\hat{\gamma} = \lambda' k_{\beta_o,\theta_o} = \Sigma \lambda_i (Y-X\beta_o)' A_i (Y-X\beta_o) \tag{4.1.5}$$

where λ is any solution of $K_{\theta_o} \lambda = f$.

Theorem 4.1.1 is established by showing that

$$\text{cov}(g(Y), \hat{\gamma} | \beta_o, \theta_o) = 0$$

for all $g(Y)$ such that $E[g(Y)|\beta,\theta] = 0 \; \forall \; \beta \in R^m$, $\theta \in F$, and using the theorem on minimum variance estimation given in Rao (1973, p. 317).

Note 1: For any λ, $\lambda' k_{\beta_o,\theta_o}$ is LMVUE of its expected value which is a linear functions of θ. Thus (4.1.5) characterizes all LMVUE's of linear functions of θ at (β_o,θ_o).

Note 2: The variance of $\hat{\gamma}$ as defined in (4.1.5) is

$$V(\hat{\gamma}|\beta,\theta) = 4(\beta-\beta_o)'X'AV_\theta AX(\beta-\beta_o) + 2 \text{ tr } AV_\theta AV_\theta \tag{4.1.6}$$

where $A = \Sigma \lambda_i A_i$ with A_i computed at θ_o. The variance at (β_o,θ_o) is

$$V(\hat{\gamma}|\beta_o,\theta_o) = 2 \lambda' K_{\theta_o} \lambda = 2 f' K_{\theta_o}^{-} f \tag{4.1.7}$$

where $K_{\theta_o}^-$ is any g-inverse of K_{θ_o}.

Note 3: The BLUE of $X\beta$ at θ_o is

$$X\hat{\beta} = P_{\theta_o} Y. \tag{4.1.8}$$

Substituting $\hat{\beta}$ for β_o in (4.1.5) we have

$$\hat{\gamma}_1 = \lambda' k_{\hat{\beta},\theta_o} = Y'(MV_{\theta_o}M)^+(\Sigma\lambda_i V_i)(MV_{\theta_o}M)^+Y \tag{4.1.9}$$

where $M = I - X(X'X)^-X'$, and C^+ is Moore Penrose inverse of C (see Rao and Mitra, 1972). The statistic $\hat{\gamma}_1$ which is independent of the apriori value of β is an alternative estimator of $f'\theta$ but it may not be unbiased for $f'\theta$.

Note 4: Theorem 4.1.1 can be stated in a different form as follows. If $f'\theta$ is unbiasedly estimable then its LMVUE at β_o,θ_o is $f'\hat{\theta}$ where $\hat{\theta}$ is any solution of the consistent equation

$$K_{\theta_o}\theta = k_{\beta_o,\theta_o} \tag{4.1.10}$$

Note 5: Let θ (i.e., each component of θ) be estimable in which case K_{θ_o} is nonsingular and the solution of (4.1.10) is $\hat{\theta}_1 = K_{\theta_o}^{-1}k_{\beta_o,\theta_o}$. Let $\hat{\beta}_1$ be a solution of $X\beta = P_{\theta_o}Y$. We may use $\hat{\beta}_1,\hat{\theta}_1$ the LMVUE of β,θ as initial values and obtain second stage estimates $\hat{\theta}_2$ and $\hat{\beta}_2$ of θ and β as solutions of

$$K_{\hat{\theta}_1}\theta = k_{\hat{\beta}_1,\hat{\theta}_1}, \quad X\beta = P_{\hat{\theta}_1}Y. \tag{4.1.11}$$

The process may be repeated and if the solutions converge they satisfy the equations

$$K_\theta \theta = k_{\beta,\theta}, \quad X\beta = P_\theta Y. \tag{4.1.12}$$

The solution $(\hat\beta,\hat\theta)$ of (4.1.12) may be called IMVUE (iterated minimum variance unbiased estimator) of (β,θ). The exact properties of $(\hat\beta,\hat\theta)$ are not known.

4.2 Invariant Estimation

Let us restrict the class of estimators to invariant unbiased (IU) estimators, i.e., estimators $g(Y)$ such that

$$g(Y + X\beta) = g(Y) \;\forall\, \beta$$

$$E[g(Y)|\beta,\theta] = f'\theta \tag{4.2.1}$$

and find the locally minimum variance invariant unbiased estimator (LMVIUE).

Let

$$M = I - P, \quad P = X(X'X)^{-}X'$$

$$H_{UI}(\theta) = (\mathrm{tr}[(MV_\theta M)^{+}V_i(MV_\theta M)^{+}V_j])$$

$$= (\mathrm{tr}[V_\theta^{-1}(I-P_\theta)V_i(I-P_\theta')V_\theta^{-1}V_j])$$

$$h_I(Y,\theta) = (Y'(MV_\theta M)^{+}V_1(MV_\theta M)^{+}Y,\ldots,Y'(MV_\theta M)^{+}V_p(MV_\theta M)^{+}Y)'$$

$$= [Y'V_\theta^{-1}(I-P_\theta)V_1(I-P_\theta')V_\theta^{-1}Y,\ldots,Y'V_\theta^{-1}(I-P_\theta)V_p(I-P_\theta')V_\theta^{-1}Y]'. \tag{4.2.2}$$

<u>Theorem 4.2.1.</u>

(i) $f'\theta$ is invariantly unbiasedly estimable iff

$$f \in S(H_{UI}(\theta)) \tag{4.2.3}$$

for any choice of θ such that V_θ is nonsingular.

(ii) The LMVIUE of $f'\theta$ at θ_o is

$$\hat{\gamma} = \lambda' h_I(Y,\theta_o) \tag{4.2.4}$$

where λ is any solution of $[H_{UI}(\theta_o)]\lambda = f$.

The results of Theorem 4.2.1 are obtained by transforming the model $Y = X\beta + \varepsilon$ to a model involving the maximal invariant of Y,

$$t = B'Y = B'\varepsilon = \varepsilon_* \tag{4.2.5}$$

where $B = X^\perp$, which is independent of β, and applying Theorem 4.1.1.

Note 1: Theorem 4.2.1 can be stated in a different form as follows. If $f'\theta$ is invariantly unbiasedly estimable, then its LMVIUE at θ_o is $f'\hat{\theta}$ where $\hat{\theta}$ is a solution of

$$[H_{UI}(\theta_o)]\theta = h_I(Y,\theta_o) \tag{4.2.6}$$

where $H_{UI}(\theta)$ and $h_I(Y,\theta)$ are defined in (4.2.2).

Note 2: If θ admits invariant unbiased estimation, then as in Note 5 following Theorem 4.1.1 we may obtain IMVIUE of (β,θ) as the solution of

$$X\beta = P_\theta Y \tag{4.2.7}$$

$$[H_{UI}(\theta)]\theta = h_I(Y,\theta).$$

5. MINIMUM NORM QUADRATIC ESTIMATION (MINQE-THEORY)

5.0 MINQE-principle

In Section 4 we assumed normal distribution for the random vector Y in the linear model and obtained the LMVUE of linear functions of variance components without imposing any restriction on the estimating function. However, we found that the estimators were all quadratic. In the present section we shall not make any distributional assumptions but confine our attention to the class of quadratic estimators and lay down some principles for deriving optimum estimators.

Natural estimator. Consider a general random effects linear model

$$Y = X\beta + U_1\phi_1 + \ldots + U_p\phi_p = X\beta + U\phi$$

$$E(\phi_i) = 0, \quad E(\phi_i\phi_i') = \theta_i G_i, \quad E(\phi_i\phi_j') = 0, \; i \neq j. \tag{5.0.1}$$

When ϕ_i is known, a natural estimator of θ_i is

$$\hat{\theta}_i = \phi_i' G_i^- \phi_i \div r_i, \quad r_i = R(G_i) \tag{5.0.2}$$

which has nice properties such as uniformly minimum variance under mild conditions on the moments of ϕ_i. Then a natural estimator of $f'\theta$, a linear function of θ, is

$$f'\hat{\theta} = \Sigma(f_i/r_i)\phi_i' G_i^- \phi_i = \phi' N \phi \quad \text{(say)}. \tag{5.0.3}$$

Given a model of the form (5.0.2) and apriori values α_i of θ_i, we can by a suitable transformation of ϕ_i

$$\phi_i \rightarrow \alpha_i^{1/2} B_i \phi_i, \quad B_i B_i' = G_i.$$

Write the model in the form

$$Y = X\beta + U_1\phi_1 + \ldots + U_p\phi_p$$

where $U_i = U_i\alpha^{1/2}B_i$

$$E(\phi_i) = 0,\ E(\phi_i\phi_i') = \theta_i I_{r_i},\ E(\phi_i\phi_j') = 0 \tag{5.0.4}$$

and the apriori values of θ_i are all equal to unity. We assume that such a transformation is made before analysis of data. In terms of the model (5.0.4), a natural estimator of θ_i is $\phi_i'\phi_i/r_i$.

It is not clear how the concept of a natural estimator can be extended to the general model

$$Y = X\beta + U\phi,\ E(\phi\phi') = \theta_1F_1 + \ldots + \theta_pF_p$$

$$D(Y) = \theta_1UF_1U' + \ldots + \theta_pUF_pU' = \theta_1V_1 + \ldots + \theta_pV_p \tag{5.0.5}$$

where ϕ may not have the structure defined in (5.0.4). However, using prior values $\alpha_1, \ldots, \alpha_p$ we may transform ϕ to $F^{-1/2}\phi$ where $F = \alpha_1F_1 + \ldots + \alpha_pF_p$, and U to $UF^{1/2}$. If (5.0.5) represents the model after such a transformation of U and ϕ (θ and V_i remain unchanged) we may formally extend the concept of a natural estimator, although we may not claim any optimal properties as in the case when ϕ has the structure of a random effects model.

<u>Definition 5.0.1</u>. Under the general model (5.0.5), a natural estimator of $\gamma = f'\theta$ is

$$\hat{\gamma} = \Sigma\ \mu_i\phi'F_i\phi = \phi'N\phi \tag{5.0.6}$$

where $N = \Sigma \mu_i F_i$, $\mu = (\mu_1, \ldots, \mu_p)'$ is a solution of $H\mu = f$, $H = (\mathrm{tr}\, F_i F_j)$.

A more general definition of a natural estimator in terms of ε when the model is $Y = X\beta + \varepsilon$ without specifying any structure for ε is given in Section 5.4.

MINQE-theory. Consider the general model (5.0.5) and a quadratic estimator $\hat{\gamma} = Y'AY$ of $f'\theta$. Now

$$Y'AY = \binom{\phi}{\beta}' \begin{pmatrix} U'AU & U'AX \\ X'AU & X'AX \end{pmatrix} \binom{\phi}{\beta} \tag{5.0.7}$$

while the natural estimator is $\phi'N\phi$ as defined in (5.0.6). The difference between $Y'AY$ and $\phi'N\phi$ is

$$\binom{\phi}{\beta}' \begin{pmatrix} U'AU-N & U'AX \\ X'AU & X'AX \end{pmatrix} \binom{\phi}{\beta}. \tag{5.0.8}$$

The minimum norm quadratic estimator (MINQE) is the one obtained by minimizing an appropriately chosen norm of the matrix of the quadratic form in (5.0.8)

$$\left\| \begin{matrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{matrix} \right\| = \left\| \begin{matrix} U'AU-N & U'AX \\ X'AU & X'AX \end{matrix} \right\|. \tag{5.0.9}$$

We shall consider mainly two kinds of norms, one a simple Euclidean norm

$$\mathrm{tr}\, D_{11}D_{11} + 2\, \mathrm{tr}\, D_{12}D_{21} + \mathrm{tr}\, D_{22}D_{22} \tag{5.0.10}$$

and another a weighted Euclidean norm

$$\mathrm{tr}\, D_{11}WD_{11}W + 2\, \mathrm{tr}\, D_{12}KD_{21}W + \mathrm{tr}\, D_{22}KD_{22}K \tag{5.0.11}$$

where W and K are n.n.d. matrices. The norm (5.0.11) gives different weights to ϕ and β in the quadratic form (5.0.8).

We impose other restrictions on A (and indicate the MINQE so obtained by adding a symbol in brackets) such as Y'AY

(a) is unbiased: MINQE(U)

(b) is invariant for translation in β: MINQE(I)

(c) satisfies both (a) and (b): MINQE(U, I)

(d) is unbiased non-negative definite: MINQE(U, D)

(e) is invariant non-negative definite: MINQE(I, D), etc.

The properties of the estimator strongly depend on the norm chosen and the restrictions imposed. We also obtain a series of IMINQE's (iterated MINQE's), by repeatedly solving the MINQE equations using the solutions at any stage as prior values for transforming the model as indicated below equation (5.0.5).

5.1 MINQE(U, I)

We consider the class of invariant unbiased quadratic estimators, i.e., of the form Y'AY where A belongs to the class

$$C_{UI}^{f} = \{A: \ AX = 0, \ \ \mathrm{tr}\, AV_i = f_i, \ \ i = 1,\ldots,p\} \tag{5.1.1}$$

where X and V_i are as defined for the general model (5.0.5). We use the following notations and assumptions

$$T = (V_\alpha + XX') > 0, \quad V_\alpha = \alpha_1 V_1 + \ldots + \alpha_p V_p = UU'$$

$$P_T = X(X'T^{-1}X)^{-}X'T^{-1}, \quad R_T = (I - P_T)$$

where α is a prior value of θ.

<u>Theorem 5.1.1</u>. If C_{UI}^{f} is not empty, then under the Euclidean norm (5.0.10), the MINQE(U, I) of $f'\theta$ is

$$\hat{\gamma} = \Sigma \lambda_i Y' A_i Y, \quad A_i = T^{-1} R_T V_i R_T' T^{-1} \tag{5.1.2}$$

where $\lambda = (\lambda_1, \ldots, \lambda_p)'$ is any solution of

$$[H_{UI}(\alpha)]\lambda = f \tag{5.1.3}$$

where $H_{UI}(\alpha)$ is the matrix $(\text{tr } A_i V_j)$.

<u>Proof</u>. Under the conditions (5.1.1), the square of the Euclidean norm in (5.0.10) becomes

$$||U'AU-N||^2 = \text{tr}(U'AUU'AU) - 2 \text{ tr } NU'AU + \text{tr } NN. \tag{5.1.4}$$

But $N = \Sigma \mu_i F_i \Rightarrow \text{tr } NU'AU = \Sigma \mu_i f_i$ so that we need minimize only the expression

$$\text{tr } AV_\alpha AV_\alpha = \text{tr } ATAT \quad \text{for} \quad A \in C_{UI}^{f}. \tag{5.1.5}$$

It is easy to show that (5.1.5) is minimized at A_* such that

$$\text{tr } DTAT = 0 \;\forall\; D \in C_{UI}^{0}. \tag{5.1.6}$$

$$D \in C_{UI}^{0} \Rightarrow D = R_T' E R_T, \quad \text{tr } E R_T' V_i R_T = 0 \text{ for arbitrary } E.$$

Then

$$T A_* T = \Sigma \lambda_i R_T V_i R_T'$$

which gives the solution (5.1.3). The equation for λ is obtained by expressing the condition of unbiasedness. Note that $[H_{UI}(\alpha)]\lambda = f$ is consistent if C^f_{UI} is not empty. Also the solution (5.1.2) is independent of N.

Note 1: MINQE(U, I)'s are additive.

Note 2: An alternative expression for $\hat{\gamma}$ given in (5.1.3) is

$$\hat{\gamma} = \Sigma\, \lambda_i Y'A_iY, \quad A_i = (MV_\alpha M)^+ V_i (MV_\alpha M)^+ \tag{5.1.7}$$

where $M = I - XX^+$.

Note 3: When V_α is nonsingular, T can be replaced by V_α in Theorem 5.1.1. Then

$$\hat{\gamma} = \Sigma\, \lambda_i Y'A_iY, \quad A_i = V_\alpha^{-1} R_{V_\alpha} V_i R'_{V_\alpha} V_\alpha^{-1}. \tag{5.1.8}$$

Note 4: If Y is normally distributed, MINQE(U, I) is LMVUE of $f'\theta$ at values of θ where $\Sigma\, \theta_i V_i$ is proportional to V_α. (See Theorem 4.1.1).

Note 5: If in (5.1.4) we use the weighted Euclidean norm (5.0.11)

$$||U'AU-N||^2 = tr(U'AU-N)\, W(U'AU-N)W \tag{5.1.9}$$

where W is p.d., the solution may not be independent of N. The expression (5.1.9) can be written as

$$tr\, AGAG - 2\, tr\, AH + tr\, NWN \tag{5.1.10}$$

where $G = UWU'$ and $H = UWNWU'$. If G is nonsingular, then the minimum of (5.1.10) is attained at

$$A_* = G^{-1}(\Sigma\ \lambda_i R_G V_i R_G' + R_G H R_G')G^{-1} \tag{5.1.11}$$

where λ_i are determined from the equations

$$\text{tr}\ A_* V_i = f_i, \quad i = 1,\ldots,p.$$

Note 6: It is seen from (5.1.2) that the estimate of $f'\theta$ can be written in the form $f'\hat{\theta}$ where $\hat{\theta}$ is a solution of

$$[H_{UI}(\alpha)]\theta = h_\alpha(Y,\alpha) \tag{5.1.12}$$

where the i-th element of $h_I(Y,\alpha)$ is

$$Y'A_iY = Y'T^{-1}R_T V_i R_T' T^{-1} Y \tag{5.1.13}$$

and $H_\alpha(U,I)$ is as defined in (5.1.3). If each component of θ admits invariant unbiased estimation then $H_\alpha(U,I)$ is non-singular and the MINQE(U, I) of θ is

$$\hat{\theta} = [H_{UI}(\alpha)]^{-1} h_I(Y,\alpha) \tag{5.1.14}$$

Note 7: The computation of MINQE(U, I) of θ involves the use of α a prior value of θ. If we have no prior information on θ, there are two possibilities. We may take α as a vector with all its elements as unity. An alternative is to choose some α, compute (5.1.14), consider it (say $\hat{\theta}_1$) as an apriori value of θ and repeat the computation of (5.1.14). The second round value, say $\hat{\theta}_2$ is an appropriate estimate of θ, which may be better than $\hat{\theta}_1$ if the initial choice α is very much different from $\hat{\theta}_1$.

We may repeat the process and obtain $\hat{\theta}_3$ choosing $\hat{\theta}_2$ as an apriori value and soon. The limiting value which satisfies the equation

$$[H_{UI}(\theta)]\theta = h_I(Y, \theta) \tag{5.1.15}$$

is the IMINQE(U, I), the iterated MINQE(U, I), which is the same as IMVIUE defined in (4.2.7). It is shown in Section 6 that the equation (5.1.15) is the restricted maximum likelihood (RML) equation considered by Patterson and Thompson (1975).

5.2 MINQE(U)

We drop invariance and consider only unbiasedness, as in problems such as those mentioned by Focke and Dewess (1972) where the condition for invariance does not hold. In such problems where invariance condition is not used, it is advisable to use an apriori value β_0 of β and change Y to $Y - X\beta_0$ and β to $(\beta-\beta_0)$ and work with the transformed model in addition to the transformation indicated in (5.0.5). The class of unbiased estimators of $f'\theta$ is defined by

$$C_U^f = \{A: \ X'AX = 0, \ \ \mathrm{tr}\, AV_i = f_i, \ \ i = 1,\dots,p\} \tag{5.2.1}$$

where X and V_i are as in the general model (5.0.5).

Theorem 5.2.1. Let $V_\alpha = UU'$ be p.d. If C_U^f is not empty then the MINQE(U) under Euclidean norm (5.0.10) is

$$\hat{\gamma} = \Sigma\, \lambda_i Y'A_iY, \ \ A_i = (V_\alpha + XX')^{-1}(V_i - P_{V_\alpha} V_i P'_{V_\alpha})(V_\alpha + XX')^{-1} \tag{5.2.2}$$

where $\lambda = (\lambda_1,\dots,\lambda_p)'$ is any solution of

$$[H_U(\alpha)]\lambda = f \tag{5.2.3}$$

where $H_U(\alpha)$ is the matrix $(\text{tr } A_i V_j)$.

Proof. Under (5.0.10) we have to minimize

$$||U'AU-N||^2 + 2||U'AX||^2$$

which, using (5.2.1), reduces to

$$\text{tr } AV_\alpha AV_\alpha + 2 \text{ tr } AV_\alpha AXX' = \text{tr } AV_\alpha A(V_\alpha + 2XX') \tag{5.2.4}$$

$$= \text{tr } ATAT, \quad T = V_\alpha + XX'. \tag{5.2.5}$$

The expression (5.2.4) attains a minimum at A_* iff

$$\text{tr } DTAT = 0 \ \forall \ D \in Q_U^0. \tag{5.2.6}$$

Observing that $D \in Q_U^0 \Rightarrow D = E - P_T' E P_T$ and following the arguments of Theorem 5.1.1, the expression for A_* is obtained as in (5.2.2), where P_T is replaced by the equivalent expression P_{V_α}.

Note 1: We shall consider a few alternatives to the simple Euclidean norm. Focke and Dewess (1972) give different weights to the two terms in (5.2.3) as in (5.0.11). Choosing $W = I$ and $K = r^2$, (5.2.4) becomes

$$\text{tr } AV_\alpha AV_\alpha + 2r^2 \text{tr } AV_\alpha AXX'. \tag{5.2.7}$$

The constant r^2 determine the relative weights to be attached to β and ϕ. The solution obtained by minimizing (5.2.7) is called r-MINQE(U) which is the same as (5.2.2) with X replaced by rX.

Note 2: The iterated estimates of β and MINQE(U) of θ are solutions of the equations

$$X'V_\theta^{-1}X\beta = X'V_\theta^{-1}Y$$

$$[H_U(\theta)]\theta = h_U(Y,\theta) \tag{5.2.8}$$

where

$$h_U(Y,\theta) = (Y'A_1Y,\ldots,Y'A_pY)', \tag{5.2.9}$$

$H_U(\theta)$ and A_i are as defined in Theorem (5.2.8). The solution of (5.2.8) is represented by IMINQE(U).

5.3 ∞-MINQE(U)

In (5.2.7) we defined r-MINQE(U) which uses a weighted Euclidean norm to provide differential wieghts to β and ϕ and also suggested a translation in Y using a prior value of β. Actually we may consider a transformation which changes

$$Y \to Y - X\beta_0,\ \beta \to r^{-1}K^{-1/2}\beta$$

where β_0 and r^2K correspond to apriori mean and dispersion of β. Then the Euclidean norm of (5.0.10) becomes

$$\text{tr}\, A(V_\alpha + r^2XKX')\, A(V_\alpha + r^2XKX') \tag{5.3.1}$$

which may be minimized.

Let us denote the optimal solution in such a case by A_r and define $A_0 = \lim A_r$ as $r \to \infty$. If A_0 exists, we call the corresponding estimator $Y'A_0Y$, the ∞-MINQE(U). The following theorem due to Focke and Dewess (1972) establishes the existence of ∞-MINQE(U).

Theorem 5.3.1. Let $\mathcal{V}$ be the set of linear combinations of $V_1,\ldots,V_p$.

(i) ∞-MINQE(U) exists if C_U^f is not empty.

(ii) A_0 is the unique matrix which minimizes $\operatorname{tr} AV_\alpha AV_\alpha$ in the class $C = \{A\colon XKX'AV_\alpha M + MV_\alpha AXKX' = V_0 - PV_0P,\ V_0 \in \mathcal{V}\}$ (5.3.2)

(iii) ∞-MINQE(U) is invariant with respect to nonsingular linear transformation of the model (5.0.5).

Theorem 5.3.1 characterizes ∞-MINQE(U) but does not provide the method of calculating it. Theorem 5.3.2 gives the formula when $V = I$, from which the formula for general V can be derived by a transformation of the model in view of statement (iii) of Theorem 5.3.1.

Theorem 5.3.2. Let $G = (\operatorname{tr} MV_iMV_j)$, $B = (\operatorname{tr} MV_i(XKX')^+V_j)$ and C_U^f be not empty. If $V_\alpha = I$, the ∞-MINQE(U) of $f'\theta$ is $\hat{\gamma} = Y'A_*Y$ where

$$A_* = (XKX')^+V_0M + MV_0(XKX')^+ + MV_bM \tag{5.3.8}$$

$$V_0 = \Sigma\, a_iV_i, \quad V_b = \Sigma\, b_iV_i$$

and $a = (a_1,\ldots,a_p)'$, $b = (b_1,\ldots,b_p)'$ satisfy the equations

$$Gb + 2\,Ba = f, \quad Ga = 0. \tag{5.3.9}$$

Theorem 5.3.3. If V_α is p.d. and C_U^f is not empty, the ∞-MINQE(U) is obtained by replacing M by $(MV_\alpha M)^+$ in (5.3.8).

Proof. We consider the model $V_\alpha^{-1/2}Y = V_\alpha^{-1/2}X + V_\alpha^{-1/2}U\phi$ and apply the result of Theorem 5.3.2. On simplication the formula stated in Theorem 5.3.3 is obtained.

Note 1: It is interesting to note that ∞-MINQE(U) is the same if instead of the sequence r^2K, we consider $(\Lambda + r^2K)$ for any $\Lambda \geq 0$.

Note 2: ∞-MINQE(U) coincides with MINQE(U, I) if it exists.

5.4 MINQE Without Unbiasedness

Let us consider the linear model

$$Y = X\beta + \varepsilon, \quad E(\varepsilon\varepsilon') = \theta_1 V_1 + \ldots + \theta_p V_p = V_\theta. \tag{5.4.1}$$

Choosing a prior value α of θ, (5.4.1) can be written

$$Y = X\beta + V_\alpha^{1/2}\varepsilon_* \tag{5.4.2}$$

where $\varepsilon_* = V_\alpha^{-1/2}\varepsilon$ and $V_\alpha = \alpha_1 V_1 + \ldots + \alpha_p V_p$. Using the definition (5.0.6) with ε_* as ϕ, a natural estimator $f'\theta$ is

$$\varepsilon_*'N\varepsilon_* = \varepsilon_*'(\Sigma\, \lambda_i V_{i*})\varepsilon_*, \quad V_{i*} = V_\alpha^{-1/2} V_i V_\alpha^{-1/2} \tag{5.4.3}$$

where $\lambda = (\lambda_1, \ldots, \lambda_p)'$ is chosen such that $\varepsilon_*'N\varepsilon_*$ is unbiased for $f'\theta$, i.e., λ satisfies the equation $[H(\alpha)]\lambda = f$ where

$$H(\alpha) = (\mathrm{tr}\, V_{i*}V_{j*}) = (\mathrm{tr}\, V_\alpha^{-1} V_i V_\alpha^{-1} V_j). \tag{5.4.4}$$

It is seen that (5.4.3) is LMVUE of θ at $\theta = \alpha$ when ε is normally distributed. In most of the applications the natural estimator is independent of α, which is an ideal situation. The MINQE of $f'\theta$ is $Y'AY$ where A is chosen to minimize

$$\cdots \left\| \begin{matrix} V_\alpha^{1/2} A V_\alpha^{1/2} - N & X'AV_\alpha^{1/2} \\ V_\alpha^{1/2} AX & X'AX \end{matrix} \right\| . \qquad (5.4.5)$$

In Sections 5.1 - 5.3 we imposed the condition of unbiasedness on $Y'AY$. We withdraw this condition but consider some alternative restrictions on the symmetric matrix A as defined by the following classes.

$$C = \{A\} \qquad (5.4.6)$$

$$C_{PU} = \{A:\ X'AX = 0\} \qquad (5.4.7)$$

$$C_I = \{A:\ AX = 0\} \qquad (5.4.8)$$

It is seen that when $A \in C_{PU}$, the bias in the estimator $Y'AY$ is independent of the location parameter β, and is thus partially unbiased (PU). The MINQE's obtained subject to the restrictions (5.4.6)-(5.4.8) are represented by MINQE, MINQE(PU), MINQE(I) respectively. The concept of MINQE(I) is due to Poduri Rao and Chaubey (1978).

<u>Theorem 5.4.1</u>. Consider the model (5.0.5) and let $V_\alpha = \alpha_1 V_1 + \ldots + \alpha_p V_p$ be p.d. Further, let $W = \Sigma \lambda_i V_i$ where $\lambda = (\lambda_1, \ldots, \lambda_p)'$ satisfies the equation $[H(\alpha)]\lambda = f$, where $H(\alpha) = (\mathrm{tr}\, V_\alpha^{-1} V_i V_\alpha^{-1} V_j)$. Then under the Euclidean norm in (5.4.5), the optimal matrix A_* providing MINQE's are as follows.

$$\text{(i) MINQE:} \quad A_* = (V_\alpha + XX')^{-1} W (V_\alpha + XX')^{-1} \qquad (5.4.9)$$

$$\text{(ii) MINQE(PU):} \quad A_* = (V_\alpha + XX')^{-1} (W - P_\alpha W P_\alpha)(V_\alpha + XX')^{-1}$$

$$P_\alpha = X(X'V_\alpha X)^{-} X' V_\alpha^{-1} \qquad (5.4.10)$$

(iii) MINQE(I): $A_* = (MV_\alpha M)^+ W(MV_\alpha M)^+,$

$$= V_\alpha^{-1}(I - P_\alpha)W(I - P_\alpha)V_\alpha^{-1} \qquad (5.4.11)$$

where $M = I - X(X'X)^- X'$.

Proof. Under Euclidean norm, the square of (5.4.5) is

$$\mathrm{tr}(V_\alpha^{1/2} A V_\alpha^{1/2} - N)^2 + 2\ \mathrm{tr}(X'AV_\alpha AX) + \mathrm{tr}(X'AX)^2. \qquad (5.4.12)$$

Without any restriction on A, the minimum of (5.4.12) is attained at A_* iff

$$\mathrm{tr}(V_\alpha^{1/2} A_* V_\alpha^{1/2} - N)V_\alpha^{1/2} B V_\alpha^{1/2} + 2\ \mathrm{tr}(X'A_* V_\alpha BX) + \mathrm{tr}(X'A_* XX'BX) = 0 \qquad (5.4.13)$$

for any symmetric matrix B. Then A_* satisfies the equation

$$V_\alpha^{1/2}(V_\alpha^{1/2} A_* V_\alpha^{1/2} - N)V_\alpha^{1/2} + XX'A_* V_\alpha + V_\alpha A_* XX' + XX'A_* XX' = 0$$

or

$$(V_\alpha + XX')A_*(V_\alpha + XX') = V_\alpha^{1/2} N V_\alpha^{1/2} = \Sigma\, \lambda_i V_i = W$$

$$A_* = (V_\alpha + XX')^{-1} W (V_\alpha + XX')^{-1}$$

which is the matrix given in (5.4.9).

If A is subject to the restriction $X'AX = 0$, then (5.4.13) must hold when B is replaced by $B - P_\alpha' B P_\alpha$ where P_α is defined in (5.4.10). Then arguing as above and noting that $P_\alpha V_\alpha = V_\alpha P_\alpha'$, the equation for A_* is

$$(V_\alpha + XX')A_*(V_\alpha + XX') = \Sigma\, \lambda_i (V_i - P_\alpha V_i P_\alpha')$$

or $$A_* = (V_\alpha + XX')^{-1}(W - P_\alpha W P_\alpha')(V_\alpha + XX')^{-1}$$

which is the matrix given in (5.4.10).

If A is subject to the condition $AX = 0$, then (5.4.13) must hold when B is replaced by MBM where $M = I - P$. Then A_* satisfies the equation

$$(MV_\alpha M)A_*(MV_\alpha M) = MWM$$

or
$$\begin{aligned} A_* &= (MV_\alpha M)^+ W(MV_\alpha M)^+ \\ &= V_\alpha^{-1}(I - P_\alpha)W(I - P_\alpha')V_\alpha^{-1} \end{aligned}$$

which is the matrix given in (5.4.11).

<u>Note 1</u>: MINQÉ in (5.4.9) and MINQE(I) in (5.4.11) are automatically non-negative, while MINQE(PU) may not be.

<u>Note 2</u>: The MINQE(I) of $f'\theta$ given in (5.4.11) can be written as $f'\hat{\theta}$ where $\hat{\theta}$ is a solution of

$$[H(\alpha)]\theta = h_I(Y,\alpha) \tag{5.4.14}$$

where $H(\alpha)$ is as defined in (5.4.4) and the i-th element of $h_I(Y,\alpha)$ is

$$Y'V_\alpha^{-1}(I - P_\alpha)V_i(I - P_\alpha')V_\alpha^{-1}Y. \tag{5.4.15}$$

The equation (5.4.14) is consistent. If θ is identifiable then $H(\alpha)$ is non-singular, in which case $\hat{\theta} = [H(\alpha)]^{-1}h_\alpha(Y,I)$. This form of the solution enables us to obtain IMINQE(I), i.e., iterated

MINQE(I), by writing $\hat{\theta}_1 = [H(\alpha)]^{-1} h_\alpha(Y,I)$ and obtaining a second stage estimate $\hat{\theta}$ with α replaced by $\hat{\theta}_1$. The limiting solution, if the process converges, satisfies the equation

$$[H(\theta)]\theta = h_I(Y,\theta) \tag{5.4.16}$$

which is shown to be the maximum likelihood equation in Section 6.

5.5 MINQE(D)-Non-negative Definite Estimator

In the general variance components model, we admitted the possibility of some of the parameters being negative. But there are cases such as the random effects model where the variance components are non-negative and it may be desirable to have non-negative estimators for them. The estimators considered so far except those in Section 5.4 can assume negative values although the parametric function is non-negative. In this section we explore the possibility of obtaining unbiased quadratic estimators $\hat{\gamma} = Y'AY$ with $A \geq 0$ of parametric functions $f'\theta$ which are non-negative in $\theta \epsilon F$ for a general model. A MINQE in this class is denoted by MINQE(U, D), where D stands for non-negative definiteness of the quadratic estimator.

The following lemma characterizes the nature of the matrix A if $\hat{\gamma}$ has to be unbiased and non-negative.

<u>Lemma 5.5.1</u>. A non-negative and unbiased quadratic estimator $Y'AY$ satisfies the invariance condition, i.e., $AX = 0$.

<u>Proof</u>. Unbiasedness $\Rightarrow X'AX = 0 \Rightarrow AX = 0$ since $A \geq 0$.

In view of Lemma 5.5.1 we need only consider the class of matrices

$$C_{UD}^{f} = \{A: \; A \geq 0, \; AX = 0, \; \mathrm{tr}\, AV_i = f_i, \; i = 1,\ldots,p\}. \tag{5.5.1}$$

Further, because of invariance we can work with a transformed model

$$t = Z'Y = \varepsilon$$

$$E(t) = 0, \;\; E(tt') = \theta_1 B_1 + \ldots + \theta_p B_p \tag{5.5.2}$$

where $Z = X^{\perp}$ (with full rank say s) and $B_i = Z'V_iZ$, $i = 1,\ldots,p$. We need consider quadratic estimators $\hat{\gamma} = t'Ct$ where C belongs to the class

$$C_{UD}^{f} = \{C: \; C \geq 0, \; \mathrm{tr}\, CB_i = f_i\}. \tag{5.5.3}$$

<u>Lemma 5.5.2</u>. C_{UD}^{f} is not empty iff

$$f \in \text{convex span } \{q(b): \; b \in R^n\} \tag{5.5.4}$$

where $q(b) = (b'MV_1Mb,\ldots,b'MV_pMb)'$.

<u>Note</u>: In terms of the model (5.5.2), the condition (5.5.4) is

$$f \in \text{convex span } \{q(b), \; b \in R^s\} \tag{5.5.5}$$

where $q(b) = (b'B_1b,\ldots,b'B_pb)$.

The conditions (5.5.4) and (5.5.5) are rather complicated, but simple results can be obtained if we assume $V_1,\ldots,V_p$ to be n.n.d.

<u>Theorem 5.5.1</u>. Let $V_i \geq 0$, $i = 1,\ldots,p$, $V = \Sigma V_i$ and $V_{(i)} = V - V_i$ and B_i be as defined in (5.5.2). There exists an n.n.d. quadratic unbiased estimator of θ_j iff

$$S(B_j) \not\subset S(B_{(j)}) \Leftrightarrow S(MV_jM) \not\subset S(MV_{(j)}M)$$

$$\Leftrightarrow S(MV_{(j)}M) \not\supset S(MVM) \Leftrightarrow R(MV_{(j)}M) < R(MVM). \qquad (5.5.6)$$

<u>Note 1</u>: The condition (5.5.6) can also be expressed as

$$(I - G)V_j(I - G) \neq 0 \qquad (5.5.7)$$

where G is the projection operator onto the space generated by the columns of the compound matrix

$$(X: V_1: \ldots : V_{j-1} : V_{j+1}: \ldots : V_p). \qquad (5.5.8)$$

<u>Note 2</u>: If $S(V_1) \supset S(M)$, then $S(MV_1M) \supset S(MV_iM)$ for all i, in which case, application of Theorem 5.5.1 shows that at most θ_1 is non-negatively estimable.

<u>Note 3</u>: If $S(V_1) \supset S(M)$ and $S(V_2) \supset S(M)$, then none of the single components are non-negatively estimable.

<u>Note 4</u>: (LaMotte, 1973.) If $V_{(j)} > 0$, then θ_j is not non-negatively estimable. Further, if $V_j > 0$, then θ_i, $i \neq j$ is not non-negatively estimable.

However, let us assume that C^f_{UD} is not empty for a given f and estimate $f'\theta$ by MINQE principle. For this purpose we have to minimize

$$||A||^2 = \text{tr } AVAV \text{ when } A \in C^f_{UD}. \qquad (5.5.9)$$

This appears to be a difficult problem in the general case. Of course, if MINQE(U, I) turns out to be a non-negative estimator in any given situation it is automatically MINQE(U, D). It may also be noted that if $sp\{MV_1M,\ldots,MV_pM\}$ is a quadratic subspace with respect to $(MVM)^+$, then the MINQE(U, I) of $f'\theta$ is n.n.d. iff C^f_{UD} is not empty.

Since C^f_{UD} is a convex set, we proceed as follows to solve the problem (5.5.9). The minimum is attained at A_* iff

$$\text{tr } BVAV \geq \text{tr } A_*VA_*V \ \forall \ B\epsilon C^f_{UD} \tag{5.5.10}$$

or writing $B = A_* + D$, the condition (5.5.12) becomes

$$\text{tr } DVA_*V \geq 0 \ \forall \ D\epsilon\mathcal{D} \tag{5.5.11}$$

$$\mathcal{D} = \{D:\ DX = 0,\ \ A_* + D \geq 0,\ \ \text{tr } DV_i = 0,\ i = 1,\ldots,p\}. \tag{5.5.12}$$

A general solution for (5.5.11) cannot be explicitly written down, but the formula will be useful in examining whether any guessed solution for A_* provides a MINQE(U, D). We shall consider some special cases.

<u>Theorem 5.5.2.</u> Let $V_i \geq 0$, $i = 1,\ldots,p$, and θ_j be estimable, i.e., the condition (5.5.9) is satisfied. Then the MINQE(U, D) of θ_j is

$$\hat{\theta}_j = \frac{1}{R(A_j)} Y'A_jY,\ \ A_j = [(I - G)V_j(I - G)]^+ \tag{5.5.13}$$

where G is the projection operator onto the space generated by the columns of $(X, V_1,\ldots,V_{j-1}, V_{j+1},\ldots,V_p)$.

An alternative approach to the problem (5.5.9) based on standard methods of convex programming is provided by Pu kelsheim (1977).

We define the functional

$$g(B) = \min_{A \epsilon C^f_{UI}} (||A||^2 - \langle A,B \rangle) \tag{5.5.14}$$

where $||A||^2 = \text{tr } AVAV$ and $\langle A,B \rangle = \text{tr } AVBV$ with $V > 0$, and call the problem

$$\sup_{B \geq 0} g(B) \tag{5.5.15}$$

as the dual optimization problem.

<u>Lemma 5.5.3</u>. Let $A_* \epsilon C^f_{UD}$ and $B_* \geq 0$ be such that

$$||A_*||^2 = g(B_*) \tag{5.5.16}$$

Then:

(i) A_* and B_* are optimal solutions of (5.5.9) and (5.5.15).

(ii) $\langle A_*, B_* \rangle = 0$. (5.5.17)

<u>Note</u>: $g(B)$ is bounded above since

$$||A_*||^2 \geq g(B) \quad \text{for all } B. \tag{5.5.18}$$

For obtaining a satisfactory solution to the problem (5.5.9) we need an explicit expression for $g(B)$. We obtain this in terms of $\hat{A}$ where $Y'\hat{A}Y$ is the MINQE(U, I) of $f'\theta$. Let us note that any matrix $B(=B')$ can be decomposed in terms of symmetric matrices

$$B = B^0 + (B - B^0)$$

such that $B^0 \epsilon C^0_{UI}$ and $\langle B^0, B-B^0 \rangle = 0$. The matrix B^0 is simply the projection of B onto the subspace C^0_{UI} in the space of symmetric matrices with inner product $\langle \cdot,\cdot \rangle$ as defined in (5.5.15). We note that by construction $\hat{A}$ is such that

$$\langle \hat{A}, B^0 \rangle = 0 \text{ for any given } B. \tag{5.5.19}$$

<u>Theorem 5.5.3</u>. Let $Y'\hat{A}Y$ be MINQE(U, I) of $f'\theta$ and C^f_{UD} be not empty. Then:

(i) $$g(B) = ||\hat{A}||^2 - \langle \hat{A},B \rangle - \frac{1}{4}\, ||B^0||^2 \tag{5.5.20}$$

(ii) B_* is optimal [i.e., maximizes g(B)] iff

$$\hat{A} + \frac{1}{2} B^0_* \geq 0, \quad \langle \hat{A} + \frac{1}{2} B^0_*, B_* \rangle = 0 \tag{5.5.21}$$

(iii) $$\hat{A}_* = A + \frac{1}{2} B^0_* \tag{5.5.22}$$

is a solution to (5.5.9), i.e., provides MINQE(U, D) of $f'\theta$ and $\langle A_*, B_* \rangle = 0$.

The results of Theorem 5.5.3 are still complicated. A sufficient condition for optimality of A_* is given in Theorem 5.5.4.

<u>Theorem 5.5.4</u>. If there exists a $B \geq 0$ such that $B^0 \neq 0$ and

$$A_* = \hat{A} - \frac{\hat{A},B}{||B^0||^2}\, B^0 \geq 0$$

then A_* is an optimal solution of (5.5.11).

<u>Proof</u>. Check optimality by the condition (5.5.16) noting that

$$B_* = -2\, \frac{\hat{A},B}{||B^0||^2}\, B.$$

6. MAXIMUM LIKELIHOOD ESTIMATION

6.1 The General Model

We consider the general GM model

$$Y = X\beta + \varepsilon$$

$$E(\varepsilon\varepsilon') = \theta_1 V_1 + \ldots + \theta_p V_p = V_\theta \tag{6.1.1}$$

and discuss the maximum likelihood estimation of θ under the assumption

$$Y \sim N_n(X\beta, V_\theta), \quad \beta \varepsilon R^m, \quad \theta \varepsilon F. \tag{6.1.2}$$

We assume that V_θ is p.d. for $\forall\ \theta \varepsilon F$.

Harville (1977) has given a review of the ML estimation of θ describing the contributions made by Anderson (1973), Hartley and Rao (1967), Henderson (1977), Patterson and Thompson (1975), Miller (1977, 1979) and others. We discuss these methods and make some additional comments.

The log likelihood of the unknown parameters (β, θ) is proportional to

$$\ell(\beta, \theta, Y) = -\log|V_\theta| - (Y - X\beta)'\, V_\theta^{-1}(Y - X\beta). \tag{6.1.3}$$

The proper ML estimator of (β, θ) is a value $(\hat{\beta}, \hat{\theta})$ such that

$$\ell(\hat{\beta}, \hat{\theta}, Y) = \sup_{\beta, \theta \varepsilon F} \ell(\beta, \theta, Y). \tag{6.1.4}$$

Such an estimator does not exist in the important case considered by Focke and Dewess (1972). In the simple version of their problem there are two random variables

$$Y_1 = \mu + \varepsilon_1, \quad Y_2 = \mu + \varepsilon_2$$

$$E(\varepsilon_1^2) = \sigma_1^2, \quad E(\varepsilon_2^2) = \sigma_2^2, \quad E(\varepsilon_1\varepsilon_2) = 0. \tag{6.1.5}$$

The likelihood based on Y_1 and Y_2 is

$$-\log \sigma_1 - \log \sigma_2 - \frac{(Y_1-\mu)^2}{2\sigma_1^2} - \frac{(Y_2-\mu)^2}{2\sigma_2^2} \tag{6.1.6}$$

which can be made arbitrarily large by choosing $\mu = Y_1$ and letting $\sigma_1 \to 0$, so that no proper MLE exists. The ML equations obtained equating the derivatives of (6.1.6) to zero are

$$\sigma_1^2 = (Y_1-\mu)^2, \quad \sigma_2^2 = (Y_2-\mu)^2, \quad \mu\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right) = \frac{Y_1}{\sigma_1^2} + \frac{Y_2}{\sigma_2^2} \tag{6.1.7}$$

which imply $\sigma_1 = \sigma_2$. Thus the ML approach fails to provide acceptable estimators. However, in the example (6.1.5), all the parameters are identifiable and MINQE(U) of σ_1^2 and σ_2^2 exist. A similar problem arises in estimating σ_1^2 and σ_2^2 in the model $Y = X\beta + X\gamma + \varepsilon$ where $E(\gamma\gamma') = \sigma_1^2 I_m$, $E(\varepsilon\varepsilon') = \sigma_0^2 I_n$ and $E(\varepsilon\gamma') = 0$.

It is well known that ML estimators of variance components are heavily biased in general and in some situations considered by Neyman and Scott (1948), they are not even consistent. In such cases, the use of ML estimators for drawing inferences on individual parameters may lead to gross errors, unless the exact distribution of the ML estimators is known. These drawbacks and the computational difficulties involved in obtaining the ML estimators make the ML method less attractive for practical applications.

6.2 Maximum Likelihood Equations

For $\theta \epsilon F$ such that $V_\theta > 0$ (i.e., p.d.), the likelihood of (β,θ) is

$$\ell(\beta,\theta,Y) = -\log|V_\theta| - (Y-X\beta)'V_\theta^{-1}(Y-X\beta). \tag{6.2.1}$$

Taking derivatives of (6.2.1) w.r.t. to β and θ_i and equating them to zero we get the ML equations

$$X'V_\theta^{-1}X\beta = X'V_\theta^{-1}Y \tag{6.2.2}$$

$$\mathrm{tr}\, V_\theta^{-1}V_i = (Y-X\beta)'V_\theta^{-1}V_iV_\theta^{-1}(Y-X\beta) \tag{6.2.3}$$

$$i = 1,\ldots,p.$$

Substituting for β in (6.2.3) from (6.2.2), the equations become

$$X\beta = P_\theta Y, \quad P_\theta = X(X'V_\theta^{-1}X)^{-}X'V_\theta^{-1} \tag{6.2.4}$$

$$[H(\theta)]\theta = h_I(Y,\theta) \tag{6.2.5}$$

where $H(\theta) = (\mathrm{tr}\, V_\theta^{-1}V_iV_\theta^{-1}V_j)$ is the matrix defined in (5.4.4) and the i-th element of $h_I(Y,\theta)$ is

$$Y'(I - P_\theta)'V_\theta^{-1}V_iV_\theta^{-1}(I - P_\theta)Y \tag{6.2.6}$$

which is the same as the expression defined in (5.4.15).

We make a few comments on the equations (6.2.4) and (6.2.5).

(i) The ML equation (6.2.5) is the same as that for IMINQE(I) given in (5.4.15).

(ii) The original likelihood equation (6.2.3) is unbiased while the equation (6.2.5) which provides a direct estimate of θ

is not so in the sense

$$E[h_I(Y,\theta)] \neq [H(\theta)]\theta. \tag{6.2.6}$$

An alternative to the equation (6.2.5) is the one obtained by equating $h_I(Y,\theta)$ to its expectation

$$h_I(Y,\theta) = E[h_I(Y,\theta)] = [H_{UI}(\theta)]\theta \tag{6.2.7}$$

which is the restricted ML (RML) equation suggested by Patterson and Thompson (1975).

(iii) There may be no solution to (6.2.5) in the admissible set F to which θ belongs. This may happen when the supremum of the likelihood is attained at a boundary point of F.

(iv) It is interesting to note that the ML estimate of θ is invariant for translation of Y by $X\alpha$ for any α, i.e., the MLE is a function of the maximal invariant $B'Y$ of Y where $B = X^{\perp}$.

Suppose θ in the model (6.1.1) is identifiable on the basis of distribution of Y in the sense: $\theta_1 V_1 + \ldots + \theta_p V_p = \theta_1' V_1 + \ldots + \theta_p' V_p \Leftrightarrow \theta_i - \theta_i' = 0$ for all i, i.e., V_i are linearly independent. But it may happen, as in the case of the example of Focke and Dewess (1972), that θ is no longer identifiable when we consider only the distribution of $B'Y$, the maximal invariant of Y. Such a situation arises when $B'V_iB$ are linearly dependent while V_i are not. In such cases the ML method is not applicable while MINQE(U) developed in Section 5.2 can be used. Thus, the invariance property of MLE limits the scope of application of the ML method.

(v) **Computational algorithms:** The equation (6.2.5) for the estimation of θ is, in general, very complicated and no closed form solution is possible. One has to adopt iterative procedures. Harville (1977) has reviewed some of the existing methods.

(a) If $\hat{\theta}_k$ is the k-th approximation to the solution of (6.2.5), then the (k+1)-th approximation is

$$\hat{\theta}_{k+1} = [H(\hat{\theta}_k)]^{-1} h_I(Y, \hat{\theta}_k) \tag{6.2.8}$$

as suggested for IMINQE(I), provided θ is identifiable. ~~on the basis of the maximal invariant of Y~~. Otherwise, the H matrix in (6.2.5) is not invertible. Iterative procedure of the type (6.2.8) is mentioned by Anderson (1973), Harville (1969), LaMotte (1973) and Rao (1972) in different contexts. However, it is not known whether the procedure (6.2.8) converges and provides a solution at which supremum of the likelihood is attained.

(b) Hartley and Rao (1967), Henderson (1977) and Harville (1977) proposed algorithms suitable for the special case when one of the V_i is an identity matrix (or at least non-singular). An extension of their method for the general case is to obtain the (k+1)-th approximation of the i-th component of θ as

$$\hat{\theta}_{i,k+1} = \hat{\theta}_{ik} \frac{Y'(I-P_{\hat{\theta}_k})' V_{\hat{\theta}_k}^{-1} V_i V_{\hat{\theta}_k}^{-1} (I-P_{\hat{\theta}_k}) Y}{\operatorname{tr} V_{\hat{\theta}_k}^{-1} V_i} \tag{6.2.9}$$

$i = 1, \ldots, p.$

In the special case when V_i are non-negative definite and the initial θ_i are chosen as non-negative, the successive approximations of θ_i using the algorithm (6.2.9) stay non-negative. This may be a "good property" of the algorithm, but it is not clear what happens when the likelihood equation (6.2.5) does not have a solution in the admissible region.

(c) Hemmerle and Hartley (1973) and Goodnight and Hemmerle (1978) developed the method of W transformation for solving the ML equations. Miller (1979) has given a different approach. Possibilities of using the variable-metric algorithms of Davidson-Fletcher-Powell described by Powell (1970) are mentioned by Harville (1977). As it stands, further research is necessary for finding a satisfactory method of solving the equation (6.2.5) and ensuring that the solution provides a maximum of the likelihood.

6.3 Restricted Maximum Likelihood Equation

As observed earlier the ML equation (6.2.5) is not unbiased, i.e.,

$$E[h_I(Y,\theta] \neq [H(\theta)]\theta. \tag{6.3.1}$$

If we replace the equation (6.2.5) by

$$\begin{aligned} h_I(Y,\theta) &= E[h_I(Y,\theta)] \\ &= [H_{UI}(\theta)]\theta \end{aligned} \tag{6.3.2}$$

we obtain the IMINQE(U,I) defined in (5.1.14), which is the same as IMVIUE defined in (4.2.7).

The equation (6.3.2) is obtained by Patterson and Thompson (1975) by maximizing the likelihood of θ based on T'Y, where T is any choice of $X^{\perp}$, which is the maximal invariant of Y. Now

$$\ell(\theta, T'Y) = -\log|T'V_\theta T| - Y'T(T'V_\theta T)^{-1}T'Y. \tag{6.3.3}$$

Differentiating (6.3.3) w.r.t. θ_i we obtain the RML (restricted ML) equation

$$tr(T(T'V_\theta T)^{-1}T'V_i) = Y'T(T'V_\theta T)^{-1}T'V_i T(T'V_\theta T)^{-1}T'Y \tag{6.3.4}$$

$$i = 1,\dots,p.$$

Using the identity (Rao, 1973, p. 77)

$$T(T'V_\theta T)^{-1}T' = V_\theta^{-1} - V_\theta^{-1}X(X'V_\theta^{-1}X)X'V_\theta^{-1}$$

$$= V_\theta^{-1}(I - P_\theta) \tag{6.3.5}$$

the equation (6.3.4) becomes

$$tr(V_\theta^{-1}(I-P_\theta)V_i) = Y'V_\theta^{-1}(I-P_\theta)V_i(I-P_\theta')V_\theta^{-1}Y \tag{6.3.6}$$

$$i = 1,\dots,p$$

which is independent of the choice of $T = X^\perp$ used in the construction of the maximal invariant of Y. It is easy to see that (6.3.6) can be written as

$$[H_{UI}(\theta)]\theta = h_I(Y,\theta) \tag{6.3.7}$$

which is the equation (6.3.2).

(i) Both ML and RML estimates depend on the maximal invariant $T'Y$ of Y. Both the methods are not applicable when θ is not identifiable on the basis of $T'Y$.

(ii) The bias in RMLE may not be as heavy as in MLE and may be more useful as point estimators.

(iii) The solution of (6.3.7) may not lie in the admissible set of θ as in the case of the ML equation.

(iv) If $\hat{\theta}_k$ is the k-th approximation, then the (k+1)-th approximation can be obtained as

$$\hat{\theta}_{k+1} = [H_{\hat{\theta}_k}(U,I)]^{-1} h_{\hat{\theta}_k}(Y,I). \tag{6.3.8}$$

It is not known whether the process converges and yields a solution which maximizes the marginal likelihood.

(v) Another algorithm for RMLE similar to (6.2.9) is to compute the (k+1)-th approximation to the i-th component of θ as

$$\hat{\theta}_{i,k+1} = \hat{\theta}_{i,k} \frac{Y'(I-P_{\hat{\theta}_k})V_{\hat{\theta}_k}^{-1}V_i V_{\hat{\theta}_k}^{-1}(I-P'_{\hat{\theta}_k})Y}{\operatorname{tr} V_{\hat{\theta}_k}^{-1}(I-P_{\hat{\theta}_k})V_i} . \tag{6.3.9}$$

It is seen that both ML and RML estimators can be obtained as iterated MINQE's, MLE being IMINQE(I) defined in (5.4.16) and RMLE being IMINQE(U, I) defined in (5.1.14). There are other iterated MINQE's which can be used in cases where ML and RML methods are not applicable.

It has been remarked by various authors that MINQE involves heavy computations, requiring the inversion of large matrices. This argument is put forward against the use of MINQE. These authors overlook the fact that inversion of large matrices depend on the inversion of smaller order matrices in special cases. For instance, if V_θ is of the form $(I + UDU')$, then it is well known that

$$V_\theta^{-1} = I - U(U'U + D^{-1})^{-1}U' \tag{6.3.10}$$

which can be used to compute V_θ^{-1} if the matrix $(U'U + D^{-1})$ is comparatively of a smaller order than V_θ. It may be noted that the computational complexity is of the same order for MINQE and MLE, RMLE.

When we have good prior information MINQE's should be better than MLE's or RMLE's.

References

Anderson, R. L. (1975). Designs and estimators for variance components. A Survey of Statistical Design and Linear Models, Ed. J. N. Srivastava, pp. 1-29.

Anderson, R. L. and Crump, P. P. (1967). Comparisons of designs and estimation procedures for estimating parameters in a two stage nested process. Technometrics, 9, 499-516.

Anderson, T. W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. Ann. Statist., 1, 135-141.

Brownlee, K. A. (1953). Industrial Experimentation, Chemical Publishing Co.,

Cochran, W. G. (1939). The use of the analysis of variance in enumeration by sampling. J. Amer. Statist. Assoc., 34, 492-510.

Fairfield Smith, H. (1936). A discriminant function for plant selection. Ann. Eugen. (London) 7, 240-260.

Fisk, P. R. (1967). Models of the second kind in regression analysis. J. Roy. Statist. Soc., B 29, 235-244.

Focke, J. and Dewess, G. (1972). Uber die Schätzmethode MINQUE von C. R. Rao and ihre Verallgemeinerung. Math. Operationforsch. u. Statist., 3, 129-143.

Fuller, W. A. and Rao, J. N. K. (1978). Estimation for a linear regression model with unknown diagonal covariance matrix. Ann. Statist., 6, 1149-1158.

Goodnight, J. H. and Hemmerle, W. J. (1978). A simplified algorithm for the W-transformation in variance component estimation. SAS Tech. Rept. R-104, Raleigh, N.C..

Hartley, H. O. and Rao, J. N. K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. Biometrika, 54, 93-108.

Harville, David A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. J. Amer. Statist. Assoc., 72, 320-340.

Hemmerle, W. J. and Hartley, H. O. (1973). Computing maximum likelihood estimates for the mixed AOV model using the W-transformation. Technometrics, 15, 819-831.

Henderson, C. R. (1950). Estimation of genetic parameters (Abstract). Ann. Math. Statist., 21, 309-310.

Henderson, C. R. (1953). Estimation of variance and covariance components. Biometrics, 9, 226-252.

Henderson, C. R. (1977). Prediction of future records. Proc. Int. Conf. on Quantitative Genetics, 616-638.

Hildreth, C. and Houck, J. P. (1968). Some estimators for a linear model with random coefficients. J. Amer. Statist. Assoc., 63, 584-595.

Kleffe, J. (1975). Quadratische Bayes-Schätzungen für Lineare Parameter Der Kuvarianzmatrix imm Gemishten Linearen Modellen. Dissertation, Humboldt Univ., Berlin.

Kleffe, J. (1976). Best quadratic unbiased estimators for variance components in mixed linear models. Sankhya B, 38, 179-186.

Kleffe, J. (1977a). Invariant methods for estimating variance components in mixed linear models. Math. Operationforsch. Statist., 8, 233-250.

Kleffe, J. (1977b). A note on ∞-MINQUE in variance covariance components models. Math. Operationforsch Statist., 8, 337-343.

Kleffe, J. (1978). Simultaneous estimation of expectation and covariance matrix in linear models. Math. Oper. Statist. Ser. Statistics, 9, 443-478.

Kleffe, J. (1979). C. R. Rao's MINQUE for replicated and multivariate observations. Tech. Rept. Zimm der AdW dur DDR, Berlin.

Krishnaiah. P. R. and Lee, Jack C. (1974). On covariance structures. Sankhya, 38A, 357-371.

LaMotte Lynn R. (1973). Quadratic estimation of variance components. Biometrics, 29, 311-330.

Miller, J. J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of analysis of variance. Ann. Statist., 5, 746-762.

Miller, J. J. (1979). Maximum likelihood estimation of variance components - a Monte Carlo Study. J. Stat. Comp. and Simmulation, 8, 175-190.

Neyman, J. and Scott, E. (1948). Consistent estimators based on partially consistent observations. Econometrica, 16, 1-32.

Olsen, A., Se ly, J. and Birkes, D. (1976). Invariant quadratic unbiased estimation for two variance components. Ann. Statist., 4, 878-890.

Panse, V. G. (1946). An application of discriminant function for selection in poultry. J. Genetics (London), 47, 242-253.

Patterson, H. D. and Thompson, R. (1975). Maximum likelihood estimation of components of variance. Proc. of 8-th International Biometric Conference, 197-207.

Pincus, R. (1974). Estimability of parameters of the covariance matrix and variance components. Math. Oper. Statist., 5, 245-248.

Powell, M. J. D. (1970). A survey of numerical methods for unconstrained optimization. SIAM Review, 12, 79-97.

Pukelsheim, F. (1977). Linear models and convex programs: Unbiased non-negative estimation in variance component models. Tech. Rep. 104, Stanford University.

Pukelsheim, F. (1978). Examples for unbiased non-negative estimation in variance component models. Tech. Rep. 113, Stanford University.

Rao, J. N. K. (1973). On the estimation of heteroscedastic variances. Biometrics, 29, 11-24.

Rao, J. N. K. and Subrahmaniam, K. (1971). Combining independent estimators and estimation in linear regression with unequal variances. Biometrics, 27, 971-990.

Rao, P. S. R. S. and Chaubey, Y. P. (1978). Three modifications of the principle of the MINQUE. Commun. Statist. Meth., A7, 767-778.

Rao, C. Radhakrishna (1947). General methods of analysis for incomplete block designs. J. Amer. Statist. Assoc., 42, 541-561.

Rao, C. Radhakrishna (1953). Discriminant function for genetic differentiation and selection. Sankhya, 12, 229-246.

Rao, C. Radhakrishna (1965). The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. Biometrics, 52, 447-458.

Rao, C. Radhakrishna (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. Proc. Fifth Berkeley Symposium, Vol. 1, 355-372.

Rao, C. Radhakrishna (1970). Estimation of heteroscedestic variances in linear models. J. Amer. Statist. Assoc., 65, 161-172.

Rao, C. Radhakrishna (1971a). Estimation of variance and covariance components. J. Multivar. Anal., 1, 257-275.

Rao, C. Radhakrishna (1971b). Minimum variance quadratic unbiased estimation of variance components. J. Multivar. Anal. 1, 445-456.

Rao, C. Radhakrishna (1972a). Estimation of variance and covariance components in linear models. J. Amer. Stat. Assoc., 67, 112-115.

Rao, C. Radhakrishna (1973). Linear Statistical Inference and Its Applications. Second Edition. John Wiley, New York.

Rao, C. Radhakrishna (1979). Estimation of variance components - MINQE theory and its relation to ML and MML estimation (in press).

Rao, C. Radhakrishna (1972) and Mitra, S. K. (1972). Generalized Inverse of Matrices and Its Applications. John Wiley, New York.

Searle, S. R. (1968). Another look at Henderson's method of estimating variance components. Biometrics, 24, 749-788.

Searle, S. R. (1971). Topics in variance component estimation. Biometrics, 27, 1-76.

Seely, J. (1970). Linear spaces and unbiased estimation - application to mixed linear model. Ann. Math. Statist., 42, 710-721.

Seely, J. (1975). An example of inadmissible analysis of variance estimator for a variance component. Biometrika, 62, 689-690.

Sinha, B. K. and Wieand, H. S. (1977). MINQUE's of variance and covariance components of certain covariance structures. Indian Statistical Institute. Tech. Rep. 28/77.

Spjotvoll, E. (1977). Random coefficients regression models. A review. Math. Oper. Statist., Ser. Statist., 8, 69-93.

Swamy, P. A. B. (1971). Statistical Inference in Random Coefficients - Regression Models. Springer-Verlag.

Yates, F. (1940). The recovery of inter-block information in balanced incomplete block designs. Ann. Eugenics (London), 10, 317-325.

Yates F. and Zacopancy, I. (1935). The estimation of the efficiency of sampling with special reference to sampling for yield in cereal experiments. J. Agri. Sc., 25, 545-577.

SECURITY CLASSIFI[illegible] OF THIS PAGE (When Data Entered)

RE[illegible] DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER AFOSR-TR-79-0974	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ESTIMATION OF VARIANCE COMPONENTS		5. TYPE OF REPORT & PERIOD COVERED Interim
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) C.R. RAO Jurgen Kleffe		8. CONTRACT OR GRANT NUMBER(s) F49620-79-C-0161
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Pittsburgh Dept. of Mathematics & Statistics Pittsburgh, PA 15160		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB, Washington, DC 20332		12. REPORT DATE 1979
		13. NUMBER OF PAGES 61
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Estimation, components of variance, MINQE, maximum likelihood (ML), Restricted maximum likelihood (RML), minimum variance.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The paper describes a number of methods for estimating variance components in a general linear model. Explicit expressions are obtained for locally minimum variance unbiased estimators with and without the invariance condition The principle of MINQE is described and a series of MINQE estimators satisfying one or more of the conditions - unbiasedness, invariance, non-negative definiteness - are derived. Corresponding to each MINQE estimator, next page

DD FORM 1473 1 JAN 73

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. Abstract continued.

iterated MINQE (IMINQE) is defined. It is shown that the ML (maximum kelihood) estimator is IMINQE satisfying the invariance condition and the L (restricted maximum likelihood estimator) is IMINQE satisfying both the variance and unbiasedness conditions. Some comments are made on the merical algorithms for computing MINQ, ML and RML estimators.

20. Abstract continued.

an iterated MINQE (IMINQE) is defined. It is shown that the ML (maximum likelihood) estimator is IMINQE satisfying the invariance condition and the RML (restricted maximum likelihood estimator) is IMINQE satisfying both the invariance and unbiasedness conditions. Some comments are made on the numerical algorithms for computing MINQ, ML and RML estimators.